## Data Analysis SIG Meeting Minutes

| Date, Time & Location: | June 4, 2004 2:00 – 3:00 EDT |
|---|---|
| Attendees: | Craig Street – Penn<br>James Lyons-Weiler – Pittsburg<br>Robert Clarke – Georgetown<br>Joseph Wang – Georgetwon<br>Antai Wong – Georgetown<br>Judith Goldberg – New York<br>David Kane – NCI/SRA<br>James Kepner – Roswell Park Cancer Institute<br>William Greco - Roswell Park Cancer Institute<br>Jihnhee Yu - Roswell Park Cancer Institute<br>Deborah Driscoll - Roswell Park Cancer Institute<br>Simon Lin – Duke<br>Patrick McConnell – Duke<br>Steve Marron – Lineberger<br>Ann Oberg – Mayo<br>Susan Geyer - Mayo<br>Richard Rauscher – Prentis-Karmanos<br>Edith Zang – Institute for Cancer Prevention<br>Naveen Vinukonda - Institute for Cancer Prevention<br>Moe Tika - Institute for Cancer Prevention<br>Brian Pittman – Institute for Cancer Prevention<br>Emily Chung - Institute for Cancer Prevention<br>Lianhong Tang - Vanderbilt<br>Pablo Tamayo – MIT/Broad Institute<br>Claire Zhu – BAH<br>Juli Klemm - BAH |
| Introduction: | **Roll-call, open meeting, review meeting goals**<br><br>- Introduction of new member: MIT/Broad Institute<br>- Review of last meeting<br>- Review mission statement<br>- Review Developer/Adopter activities |
| Review Discussions: | **Review discussion of last meeting**<br><br>- The group discussed challenges that are faced by experimentalists when analyzing their functional genomics data, the limitations of training in statistical data analysis and the scarcity of biostatisticians to collaborate with.<br>- The group identified training and documentation as important issues not only within this SIG, but also across the board in caBIG.<br>- A position paper is being put together to outline some the major challenges and issues and to suggest recommendations that will guide future activities. Louise Showe of Wistar is taking a lead on this and James Lyons-Weiler volunteered to collaborate. It is expected that the white paper will capture some of the issues identified from the SIG meetings. |

## Review of Mission Statement

- The purpose of the mission statement is to identify needs and opportunities and to facilitate communications within and outside the caBIG community.

- Group review of the mission statement.

    o Identify key categories of end users in the Opportunities/Needs section (Patrick McConnell).

    o The descriptions of user communities in the mission statement were not sufficient, as biostatisticians belong to both categories (Judy Goldberg).

    o A third community could be included that represents professional and commercial software venders, the idea being that commercial vendors produce high-quality, validated tools that are of great use to the community (Jim Kepner).

        ▪ The group discussed the issues around commercial vendors' involvement in caBIG. David Kane reminded the group that at the kickoff meeting, it was conveyed that ultimate goal of caBIG is interoperability and sharing and that the open source requirements are for caBIG-funded development. It was agreed that commercial involvement should be explored further within this SIG.

    o The end-user communities are very diverse, and resources that tailor towards specific groups of users can be very useful. It may be useful for the group to create a compendium of data analysis resources that are organized by data types (Craig Street)

    o This SIG may also focus on methodologies that combine a set of tools to form a chain of methods that can be reproduced and distributed (Pablo Tomayo, MIT/Broad)

## Review of Developer/Adopter activities

Juli has been contacting centers individually and discussing resources and timelines with each centers as part of the matchmaking process. The following updates represent the current thinking and may evolve in the future.

- Dartmouth will adapt Q5 to the caBIG architecture. Interested adopters are The Institute for Cancer Prevention and New York.

- Georgetown will adapt VISDA to the caBIG. Wistar is the likely adopter for this tool.

- UC San Francisco will adapt Magellan to the caBIG architecture. Penn will be the adopter for this tool.

- Lineberger will adapt DWD to the caBIG architecture. Wistar is the likely adopter for this tool.

- Fox Chase will adapt FGDP to the caBIG architecture. The Institute for Cancer Prevention will be the adopter for this tool.

- Duke will develop JavaR for caBIG. Oregon Health will be the adopter for this project.

- Duke will develop RProteomics for caBIG. Oregon Health and Penn are interested adopters.

- caGEDA from Pittsburg is an unfunded development effort - publication about this
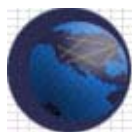
# Integrative Cancer Research Special Interest Group Teleconference

|  | tool has just been released. The Institute for Cancer Prevention will adopt this tool as an unfunded effort. |
|---|---|
|  | **Overview of Gene Pattern** |
|  | - The group was given an introduction to the Gene Pattern Tool from MIT/Broad (Pablo Tomayo).  GenePattern is an analysis platform that supports the integration of analysis tools from disparate sources. It enables the user to add new tools but requires minimum software engineering.  GenePattern can be downloaded at: **http://www.broad.mit.edu/cancer/software/genepattern/** |
| **Future Activities** | **Current projects**<br><br>- The group agreed to use upcoming meetings for presentations/demos of the tools within this SIG.<br><br>    o ~ 20 min presentation/demo on each tool, followed by ~ 10 min discussions.<br><br>    o Can be PowerPoint format or live demo through Centra.<br><br>    o For next meeting, which is on July 2nd, Jim Lyons-Weiler and Steve Marron will present caGEDA and DWD, respectively. Juli will follow up on format and resources.<br><br>In the future, this meeting will be used to update on ongoing projects, and to resolve issues and problems. |
| **Other Items Discussed** | - A question was raised regarding how data types would be normalized across applications (Steve Enkemann).  Juli noted that there is a presentation of Common Data Elements at the upcoming ICR WS/WG meeting where this type of issue will be addressed.<br><br>- There was a discussion on what should be the focus of ICR efforts - selecting the best tool or validating tools? It was agreed that the focus should be on validating tools. A compendium of literature focused on tool comparison would be a useful resource for this group.<br><br>    o Naveen provided the following link to microarray data analysis literature: **http://www.nslij-genetics.org/microarray/**<br><br>    o Jim Lyons-Weiler provided a link to the compendium of analysis tools he has created: **http://bioinformatics.upmc.edu/Help/MicroarrayReferences.html**<br><br>- Time for next meeting (July 2nd) will be changed to 12:00 PM EDT, due to the July 4th holiday weekend |

| Action Items: | Name Responsible | Action Item | Date Due | Notes |
|---|---|---|---|---|
|  | Juli Klemm | Distribute meeting minutes | 6/11/04 |  |
|  | Jim Lyons-Weiler and Steve Marron | Present caGEDA and DWD at next month's meeting | 7/2/04 |  |
|  | Juli Klemm | Follow up with James and Steve on presentations/demo | 6/21/04 |  |
|  | Juli Klemm | Follow up with Louise Show and Jim Lyons-Weiler re: SIG whitepaper | 6/14/04 |  |

**Integrative Cancer Research Special Interest Group Teleconference**

| | Juli Klemm | Update and redistribute mission statement | 6/11/04 | |
|---|---|---|---|---|